



Article Short-Term Load Probabilistic Forecasting Based on Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise Reconstruction and Salp Swarm Algorithm

Tianyu Hu, Mengran Zhou *, Kai Bian, Wenhao Lai and Ziwei Zhu

School of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan 232001, China; alexhuty153@163.com (T.H.); kbian92@163.com (K.B.); whlai@aust.edu.cn (W.L.); zwzhu0039@163.com (Z.Z.)

* Correspondence: mrzhou@aust.edu.cn

Abstract: Short-term load forecasting is an important part of load forecasting, which is of great significance to the optimal power flow and power supply guarantee of the power system. In this paper, we proposed the load series reconstruction method combined improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) with sample entropy (S_E). The load series is decomposed by ICEEMDAN and is reconstructed into a trend component, periodic component, and random component by comparing with the sample entropy of the original series. Extreme learning machine optimized by salp swarm algorithm (SSA-ELM) is used to predict respectively, and the final prediction value is obtained by superposition of the prediction results of the three components. Then, the prediction error of the training set is divided into four load intervals according to the predicted value, and the kernel probability density is estimated to obtain the error distribution of the training set. Combining the predicted value of the prediction set with the error distribution of the corresponding load interval, the prediction load interval can be obtained. The prediction method is verified by taking the hourly load data of a region in Denmark in 2019 as an example. The final experimental results show that the proposed method has a high prediction accuracy for short-term load forecasting.

Keywords: load forecasting; load series; mode decomposition; extreme learning machine; kernel density estimation

1. Introduction

With the development of industry and the economy, the conflict between supply and demand for energy is becoming increasingly acute. Among them, electric energy is not only closely related to people's lives, but also closely related to industrial production. Therefore, the balance between the supply and demand of electric energy is of particular concern. At present, the main power generation model in the world is still coal combustion power generation, which will cause air pollution. To ensure the sustainable development of economy, countries all over the world are vigorously developing new energy [1]. With the development of electric energy conversion technology and electric energy storage technology [2,3], photovoltaic power generation, wind power generation, tidal power generation, and geothermal power generation are more and more incorporated into the power grid, which not only alleviates the energy shortage but also introduces a large number of random power flows. This poses a new severe challenge to the stability and load balance of the power grid.

In the power system incorporating a large number of new energy sources, power needs to achieve a two-way balance between supply and demand. However, due to the uncontrollability of the power generation on the supply side being affected by a variety of influencing factors, the power consumption behavior of users on the demand side also has certain randomness. The interaction between supply and demand increases more



Citation: Hu, T.; Zhou, M.; Bian, K.; Lai, W.; Zhu, Z. Short-Term Load Probabilistic Forecasting Based on Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise Reconstruction and Salp Swarm Algorithm. *Energies* **2022**, *15*, 147. https://doi.org/10.3390/ en15010147

Academic Editors: Periklis Gogas and Theophilos Papadimitriou

Received: 19 November 2021 Accepted: 21 December 2021 Published: 27 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). uncertain factors for the load flow of the system, and accurate short-term load forecasting is of great significance to ensure the balance of the power system [4]. On the other hand, since September 2021, China has notified many places to limit the power load, which has had a certain impact on the lives of some people and the production of enterprises. Therefore, accurate prediction of power load is a major demand for social development. Finally, with the construction of the smart grid [5], it is not only to improve the stability and energy utilization of the system, and reduce the power generation cost, but also an important goal. Accurate prediction of power demand in various regions is helpful to realize the economic operation of a power system [6].

Load forecasting can be divided into point forecasting [7,8] and probability forecasting [9,10] according to the forecasting results. At present, most load forecasting is mainly point forecasting of load, and the forecasting result is the single point expectation of load at a certain time in the future. Power load is nonlinear and time-varying, so point prediction is difficult to reflect the fluctuation range of load change. The estimation of some uncertain factors in power market by probabilistic prediction method is helpful to the control and stable operation of power grid [11].

According to whether the prediction object or the distribution type of prediction error presupposes, probability prediction can fall into parametric probability prediction [12] and nonparametric probability prediction [13,14]. Using the parametric methods for probability density estimation requires the object is estimated to conform to a specific distribution, which has limitations in the present situation where more and more new energy generation is being integrated into the grid. The a priori assumptions avoided by the non-parametric method and the absence of excessive human intervention make it easier to approach the actual distribution.

In most decomposition and integration models, the load series is decomposed into several components by decomposition method. Then, predicting each component, the number of models is large, and the training time is long. In order to solve this problem, we use improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) combined with sample entropy to reconstruct the load series into three parts: random component, periodic component, and trend component, which reduces the number of models. In this way, the number of prediction models can be reduced to three and the training time can be shortened. For most load forecasting, point forecasting is used, which is difficult to reflect the load variation range. We use point forecasting combined with probability forecasting to predict the load interval. The error interval of the prediction set is obtained by combining the probability distribution of the error of the training set with KDE, and the final prediction interval can be obtained by combining the predicted value of the point. Finally, under the 90% confidence interval, the prediction intervals coverage probability (PICP) reached 0.919, indicating that 91.9% of the prediction set data fell within the prediction interval. On the other hand, the prediction intervals normalized averaged width (PINAW) on the cover is 0.112, which shows that we do not improve the prediction accuracy by increasing the bandwidth. In conclusion, we can draw a conclusion that the method proposed in this paper has good prediction accuracy and has a good application prospect in the field of load probabilistic forecasting.

The rest of this paper is structured as follows. The second section introduces the current research work of load forecasting. The third section introduces the relevant methods used in this paper. The fourth section mainly introduces the realisation process of the model and evaluation indicators. The fifth section is the experimental results and analysis. The sixth section is the summary of this paper.

2. Literature Review

At present, load forecasting methods are mainly divided into traditional methods and artificial intelligence methods. Artificial intelligence methods mainly include deep learning methods represented by the short-term memory network (LSTM) [15,16] and the convolutional neural network (CNN) [17,18], and machine learning methods represented by support vector regression (SVR) [19,20] and the artificial neural network (ANN) [21,22]. The deep learning method has the characteristics of a good prediction effect and high fault tolerance to input, but the model spends a lot of time in training. At present, decomposition and integration models have made preferable effects in load forecasting and other energy forecasting fields, but these models often predict all decomposed components one by one and then superimpose the results, so the training time is usually long. In addition, there is a direct relationship between the decomposition and the prediction accuracy of the integrated model and the decomposition (EMD) [23]. The amplitude and iteration number of white noise added by ensemble empirical mode decomposition (EEMD) [24] depends on the human experience setting. When the numerical setting is not set, it may be unable to overcome the phenomenon of modal aliasing. These factors may affect the prediction results.

At present, most load forecasting still takes the determined load value as the forecasting goal. Ge et al. [25] achieved good accuracy in industrial load prediction using reinforcement learning combined with least squares support vector machines for particle swarm optimisation. Zhang et al. [26] used complete ensemble empirical mode decomposition with adaptive noise combined with support vector regression with dragonfly optimization to forecast the electric load, which also had good prediction results. Rafi et al. [27] used convolutional neural networks combined with long- and short-term memory networks to construct a prediction model for short-term electricity load forecasting and achieved good prediction reliability. Wang et al. [28] used a long- and short-term memory network to forecast short-term residential loads with consideration of weather features. Phyo et al. [29] used classification and regression tree and the deep belief network for 30-min granularity load forecasting.

On the other hand, deterministic forecasting is difficult to fully reflect the load information. Therefore, using the probability forecasting method to predict the load change range is helpful to provide strong support for the production, dispatching, operation, and other links of the power grid system.

In addition, the prediction accuracy of decomposition and the integrated model is directly related to the decomposition method, and the phenomenon of mode aliasing may occur in empirical mode decomposition. On the other hand, most decomposition and integration models build prediction models for each component. Although the prediction accuracy is high, the number of models is large and the training time is long.

In this paper, we first carry out point prediction, and then analyze the training set error to obtain the distribution of prediction error in different load intervals to realize load probability prediction. The improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) [30] effectively solves the problem of mode mixing in empirical mode decomposition (EMD) and avoids the residual noise in decomposition ensemble empirical mode decomposition (EEMD), which helps to improve the prediction accuracy of the model. Firstly, the ICEEMDAN combined with sample entropy is used to reconstruct the load series, which is decomposed into three parts-random component, periodic component, and trend component-which effectively reduces the number of prediction models and shortens the prediction time. Since the extreme learning machine (ELM) algorithm was proposed, it has achieved good results in many fields, such as fault diagnosis [31,32], coal mine safety [33], and so on. The accuracy of the prediction results can be effectively improved by using the salp swarm algorithm (SSA) to optimize the ELM. Then, the kernel density estimation method is used to analyze the training set error, obtain the probability density curve of the training set error, and then estimate the error interval of the prediction set to obtain the final interval prediction result.

3. Methods

3.1. Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN)

Improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) is an algorithm based on empirical mode decomposition (EMD) proposed by Colominas et al. [34]. ICEEMDAN can effectively solve the mode mixing problem of EMD and the residual noise problem of EEMD. The decomposition process is as follows:

(1) Calculate the local mean of $S^{(i)} = S + \lambda_0 C_1(\alpha^{(i)})$ by EMD to obtain the first-order residue R_1 and corresponding intrinsic mode function (IMF) *IMF*₁.

$$R_1 = \left(M\left(S^{(i)}\right) \right) \tag{1}$$

$$IMF_1 = S - R_1 \tag{2}$$

where $i \in \{1, 2, 3 ... M\}$, *S* is the original signal; λ is the signal-to-noise ratio; $\alpha^{(i)}$ be a realization of zero mean unit variance white noise; $C_j(\cdot)$ is the operator represents the *j*th order intrinsic mode function obtained by EMD; and $M(\cdot)$ is the operator represents the local mean of the resulting signal.

(2) Calculate the local mean of $R_1 + \lambda_1 C_2(\alpha^{(i)})$ by EMD to obtain the second-order residue R_2 and corresponding intrinsic mode function IMF_2 .

$$R_2 = \left(M \left(R_1 + \lambda_1 C_2 \left(\alpha^{(i)} \right) \right) \right)$$
(3)

$$IMF_2 = R_1 - R_2 \tag{4}$$

(3) Repeat the process until the signal cannot be decomposed.

$$R_{l} = \left(M \left(R_{l-1}(t) + \lambda_{l-1} C_{l} \left(\alpha^{(i)} \right) \right) \right)$$
(5)

$$IMF_l = R_{l-1} - R_l \tag{6}$$

where l = 2, 3, ..., L, *L* are the total numbers of IMF.

Finally, the original signal is decomposed into $S = \sum_{j=1}^{L} IMF_j + R_L$.

3.2. Sample Entropy S_E

Sample entropy (S_E) [35] is a method to measure the complexity of unstable time series. Compared with the general method, the sample entropy method does not depend on the data length and has a better consistency. The value of sample entropy is positively correlated with the degree of sequence self-similarity. The sample entropy is calculated as follows:

(1) For the time series x(i) with sample size N, the following vectors are obtained according to the order of m dimensional vectors of the time series:

$$X(i) = [x(1), x(2), \dots, x(N-m+1)]$$
(7)

where, i = 1, 2, 3 ... N - m + 1.

(2) C group optimization algorithm proposed by Mirjaln $X_m(i)$ whose distance from $X_m(j)$ is less than r in $X_m(i)$. Define this number as B_i . The ratio of B_i to the total number of vectors is denoted B_i^m .

$$d_m = [X_m(i), X_m(j)] = \max_{0 \le k \le m-1} |x(i+k) - x(j+k)|$$
(8)

$$B_i^m(r) = \frac{B_i}{N-m+1} \tag{9}$$

$$B^m(r) = \frac{\sum_{i=1}^{m} B_i^m}{N-m}$$
(10)

- (3) Increase the dimension to m + 1, and repeat the step to calculate the $B^{m+1}(r)$
- (4) Calculate sample entropy S_E .

$$S_E = -\ln\left[\frac{B^{m+1}(r)}{B^m(r)}\right]$$
(11)

3.3. Salp Swarm Algorithm (SSA)

The Salp Swarm Algorithm (SSA) is a heuristic group optimization algorithm proposed by Mirjalili et al. [36] in 2017. The SSA algorithm mimics the swarm behaviour of salp on the seabed to find the optimal parameters. In the sea, the salp group is in a chain shape; the frontmost salp is responsible for guiding the whole swarm, and the following salps are responsible for searching the global situation according to the forward direction. The specific process of the SSA is as follows:

Initialize all parameters, the number of salp is *M*, the maximum number of iterations is *I*, and [*lb*, *ub*] is the search range. *d* is the dimension of the parade target.

(1) Population initialization. SSA initializes the population by generating random numbers.

$$X_{M \times d} = lb + rand(M, d) \times (ub - lb)$$
⁽¹²⁾

- (2) Calculate the fitness of each salp. Save the salp coordinates with the highest fitness.
- (3) Calculate variable c_1 .

$$c_1 = 2e^{-\left(\frac{4i}{T}\right)^2} \tag{13}$$

In the Equation (13), *i* is the current iteration number; and *I* is the maximum iteration number.

(4) Update the first salp's position. The first is responsible for searching for food to lead the movement direction of this salp population. The update equation the position of the first salp is:

$$x_d^1 = \begin{cases} P_d + c_1((ub_d - lb_d)c_2 + lb_d), c_3 \ge 0.5\\ P_d - c_1((ub_d - lb_d)c_2 + lb_d), c_3 < 0.5 \end{cases}$$
(14)

where, x_d^1 denotes the position of the leader of the salp in *d* dimensional space; ub_d and lb_d are upper and lower bounds of *d* dimensional space, respectively. P_d is the position of food source in *d* dimensional space; c_2 and c_3 are random numbers uniformly generated within the range of [0, 1].

(5) Update the location of the follower, update the equation is:

$$x_d^m = \frac{1}{2} \left[x_d^m + x_d^{m-1} \right]$$
(15)

where, $m \ge 2$, x_d^m is the position parameter of the *m*th salp in the *d* dimensional space. Calculate the fitness of each salp. Save the salp coordinates with the highest fitness.

Update iteration number *i* = *I* + 1.
(7) If the *i* > *I*, then output the coordinates of the salp with the optimal fitness. Otherwise skip to step (3).

3.4. Extreme Learning Machine (ELM)

(6)

Extreme learning machine (ELM) [37] is proposed by Huang et al. It is a supervised learning method for a single hidden layer feedforward neural network. The input weight

matrix and hidden layer threshold of ELM are randomly generated, which has the advantages of fewer training parameters and a short training time.

The mathematical model of ELM is as follows:

$$y_i = \sum_{j=1}^{l} g(\omega_j \cdot x_i + b_j) \cdot \beta_j$$
(16)

In the Equation (19), i = 1, 2, ..., N; x_i is the input vector; y_i is the output vector; g(x) is the incentive function; ω_j is the input weight matrix; b_j is the hidden layer threshold; β_j is the output weight matrix; l is the number of hidden layer nodes; and N is the number of samples.

3.5. Kernel Density Estimation (KDE)

Kernel density estimation (KDE) [38–40] is proposed by Parzen, mainly by using differentiable kernel function to estimate the probability density function.

$$\hat{f}(x) = \frac{1}{Mw} \sum_{i=1}^{M} F\left(\frac{x-x_i}{w}\right)$$
(17)

In the formula, *M* is the number of samples; F(x) is a kernel function, which includes Normal kernel, Box kernel, Triangle kernel, Epanechnikov kernel; *w* is the window width.

4. Realisation Process and Evaluation Index

4.1. Realisation Process

Although the traditional decomposition "model and ensemble" prediction model has a good prediction effect, it also needs to establish forecasting models for all components separately, which requires a lot of training time. In this paper we reconstructed the ICEEMDAN decomposed components by combination with sample entropy and load characteristics. Specifically, the load is divided into a stochastic component, a periodic component, and a trend component. Then, the three components are predicted respectively, and the final point prediction result is obtained by superimposing the prediction results of the three components. The specific prediction process of the model is as follows:

- Decomposition of load data. ICEEMDAN is used to decompose the original load series to obtain some IMF. Then, calculate the sample entropy of the original series and each IMF.
- (2) Reconstruction of load data. The IMF with sample entropy greater than 0.5 is reconstructed as the random component, the IMF with sample entropy less than 0.04 is reconstructed as the trend component, and the remaining IMF is reconstructed as the periodic component.
- (3) Forecasting of load values. The data set contains 8760 load data. The training set and prediction set are divided according to 4:1. The first 7008 load data are used as the training set, and the remaining data are used as the prediction set. Use SSA-ELM to establish models for random component, periodic component, and trend component respectively for prediction. Take the load value two hours before the prediction time as input to obtain the prediction results of each component, and overlay the three results to get the final point prediction results. SSA searches the number of hidden layer neurons and hidden layer threshold of ELM group optimization to improve the prediction performance of ELM.
- (4) Normalisation of error data. To avoid the effect of predicted value size on the error estimates, the error values were normalised using the maximum actual value of the load in the training set.
- (5) Calculate the upper and lower limits of error. Several error intervals are divided according to the prediction results of the training set. The kernel density estimation is used to obtain the probability density function of each interval training set error.

Select the appropriate kernel function by fitting the probability density function image and real error data fitting. Combined with interval confidence, the upper and lower error limits are obtained.

(6) Obtain the final prediction interval by superimposing the load value of the prediction set with the corresponding upper and lower limits of error.

4.2. Evaluation Index

To evaluate the point prediction results of the proposed model, we use the mean absolute percentage error (MAPE), mean absolute error (MAE), and mean square error (MSE) to evaluate the accuracy of the prediction results. The equations are as follows:

$$MAPE = \frac{1}{M} \sum_{i=1}^{M} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$
(18)

MAE =
$$\frac{1}{M} \sum_{i=1}^{M} |y_i - \hat{y}_i|$$
 (19)

MSE =
$$\frac{1}{M} \sum_{i=1}^{M} |y_i - \hat{y}_i|^2$$
 (20)

In the above equations, *M* is the number of samples; y_i is the actual load value; and \hat{y}_i is the predicted load value.

To evaluate the interval prediction results, PICP and PINAW are introduced. The equations are as follows:

$$PICP = \frac{1}{M} \sum_{i=1}^{M} c_i \tag{21}$$

PINAW =
$$\frac{1}{MR} \sum_{i=1}^{M} |U_i - L_i|$$
 (22)

In the formula, *M* represents the number of samples; when the prediction result is in the interval, $c_i = 1$; when the prediction result is not in the interval, $c_i = 0$; *R* is the true value range; U_i is the upper bound of prediction; and L_i is the lower bound of prediction.

5. Experiments and Analysis

5.1. Experimental Data and Conditions

To further test the prediction performance of the model, we use the hourly load data of a region in Denmark in 2019 for verification obtained from ENTSO-E. The load value is shown in Figure 1. We can see that the load value is generally stable, and the distribution shows a trend of high, medium, and low at both ends.

Experiments were conducted on 64-bit Windows 10 using MATLAB R2018a with an i7-7700hq CPU and a GTX-1050 graphics card.

From the Figure 1, we can see that the load data at 5–7 p.m. on May 1 is 0, which may be the abnormal data caused by missing data. At 8:00 a.m. and 9:00 a.m. on November 4, the load reached the highest value of the whole year, but this value is relatively isolated. This situation also shows that the change of load is affected by many factors and has some randomness. On the whole, the fluctuation of annual load data is small, and the load at the beginning and end of the year is slightly larger in the overall trend.



Figure 1. Load value of a region in Denmark in 2019.

5.2. Selection of Mode Decomposition Method

Firstly, empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD) and improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) are used to decompose the original load series. To control the experimental variables, we set the noise weight of EEMD and ICEEMDAN to 0.2 and the number of noise additions to 50. A higher entropy value of the intrinsic mode function (IMF) means a lower autocorrelation of the IMF. The results are shown in Table 1. The sample entropy of the original series is 1.462. The higher the sample entropy, the lower the autocorrelation of the IMF series and the more complex the IMF. The sample entropy of IMF 11 and IMF 12 generated by EEMD decomposition is 0, because the sample entropy of the two IMF is less than 1×10^{-5} . The series is chaotic and random. Table 1 shows the sample entropy values and correlation coefficients for each IMF.

Table 1. Sample entropy and correlation coefficient.

Method		IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9	IMF10	IMF11	IMF12	IMF13
EMD	$CC S_E$	0.087 0.192	0.352 0.518	0.672 0.6173	0.203 0.1767	0.280 0.329	0.244 0.155	0.104 0.247	0.155 0.041	0.217 0.084	0.324 0.341	0.033		
EEMD	CC S _F	0.205 0.763	0.582 1.121	0.607 0.873	0.232 0.092	0.380 0.064	0.272 0.300	0.120 2.63×10^{-3}	0.195 3.00×10^{-3}	0.339 1.42×10^{-3}	0.153 9.56×10^{-4}	0.336 0	0.308 0	$0.191 \\ 2 \times 10^{-5}$
ICEEMDAN	сс s _E	0.193 0.729	0.511 1.123	0.618 1.059	0.195 0.108	0.342 0.082	0.212 0.041	$\substack{0.065\\ 4.30 \times 10^{-3}}$	$\begin{array}{c} 0.134 \\ 3.30 \times 10^{-3} \end{array}$	$0.350 \\ 1.66 \times 10^{-3}$	$\substack{0.019\\ 1.30 \times 10^{-3}}$			

We reconstruct the IMF with entropy > 0.5 into a random component. The IMF with 0.04 < entropy < 0.5 is reconstructed into a periodic component. IMF with entropy < 0.04 is reconstructed as a trend component. The composition of the three components under different modal decomposition methods is shown in Table 2.

Table 2. Division of three components by different decomposition methods.

Method	Random Component	Periodic Component	Trend Components
EMD	IMF1-IMF3	IMF4–IMF7	IMF8–IMF11
EEMD	IMF1–IMF3	IMF4–IMF6	IMF7–IMF13
ICEEMDAN	IMF1-IMF3	IMF4–IMF6	IMF7-IMF10

According to the division results in Table 2, we reconstructed the decomposed load series and then used the extreme learning machine (ELM) to predict the results as shown in the following Table 3. When the ELM algorithm is used for prediction, to ensure the optimal number of neurons in the hidden layer, we set a cycle, that is, the number of hidden neurons is from 1 to 100, and the optimal number of neurons is selected. The prediction

results are shown in Table 3. It can be seen that the accuracy of load series prediction after decomposition and reconstruction using ICEEMDAN algorithm is the highest, absolute percentage error (MAPE) is 2.50, mean absolute error (MAE) is 63.84, and mean square error (MSE) is 9625.20. The prediction results based on EMD decomposition and reconstruction are worse. It is possible that a mode mixing situation has occurred. Therefore, we can judge that using ICEEMDAN to reconstruct and predict the load series has good accuracy.

Table 3. Prediction results of ELM.

Method	MAPE(%)	MAE	MSE
EMD-ELM	2.60	67.23	16,393.89
EEMD-ELM	2.66	68.20	12,555.00
ICEEMDAN-ELM	2.50	63.84	9625.20

Based on the above experimental results, we choose to use ICEEMDAN combined with sample entropy reconstruction to decompose the load data. The reconstructed load data is shown in Figure 2.



Figure 2. The reconstructed load series.

Combined with Figure 2, we can see that the load value showed a downward trend from January to August, reaching the bottom of electricity consumption in August, and the load value showed an upward trend from August to December. Through the variance and standard deviation, we can find that the January, February, April, and December load values is bigger, and the June, July, August, and September load values is smaller.

Figure 2 is the three load components reconstructed by ICEEMDAN combined with sample entropy. We can see that the periodic component has obvious and stable periodicity; when the fluctuation range of the trend component is small, the load value is high at both ends and low in the middle, and the overall trend is similar to that of the original data. The series with a higher frequency of random component variation is more ambiguous, and the variation range of load value is large and random. Through the above analysis, we can conclude that the reconstructed component conforms to the characteristics of the original load data.

5.3. Prediction Performance of Different Prediction Methods

To select the best prediction algorithm, we chose the BP neural network, support vector regression, and ELM to compare. The predicted results are shown in Table 4. The experimental results are shown in Table 3. The MAPE and MAE of ICEEMDAN-ELM are greater then ICEEMDAN-BP, and MSE is smaller than that of ICEEMDAN-BP. However, the three evaluation indexes of ICEEMDAN-ELM are better than ICEEMDAN-SVR. As MSE is more sensitive to extremum, combining the three evaluations we chose ICEEMDAN-ELM.

Table 4. Prediction results of different algorithms.

Method	MAPE(%)	MAE	MSE		
ICEEMDAN-BP	2.28	58.68	9822.40		
ICEEMDAN-SVR	3.13	77.10	11,582.00		
ICEEMDAN-ELM	2.50	63.84	9625.20		

In the experimental process, we find that although ELM has the advantages of high accuracy and a fast training speed, the prediction stability is slightly poor. To further improve the prediction effect, we use the salp swarm algorithm (SAA) to optimize the number of hidden layer neurons and threshold of ELM to improve the accuracy of point prediction. After using SSA optimization, the prediction accuracy of the model has been significantly improved. It can be seen that MAPE, MAE and MSE decreased to 1.98, 50.42 and 6723.70, respectively. Figure 3 is the comparison between the prediction results of SSA-ELM and ELM. From Figure 3, we can see that SSA-ELM has a higher prediction accuracy than ELM. Therefore, we can conclude that using the SSA method to optimize the number and threshold of ELM hidden layer neurons is better than selecting only the optimal number of ELM hidden layer neurons.



Figure 3. Comparison of actual and predicted values.

5.4. Performance of Reconstructed Model and Ordinary Model

To better evaluate the three different prediction models, we use SSA-ELM to predict the load data processed by different methods. From Table 5, we can see that the prediction effect of the model combined with ICEEMDAN is better than that of the ordinary model without decomposition. On the other hand, we can see that the training time of the reconstructed model is 127.78 s, which is significantly lower than that of the decomposed model. Considering the prediction accuracy, the number of models, and training time, we believe that the overall performance of the reconstructed model is better.

Method	MAPE(%)	MAE	MSE	Traing Time (s)
Reconstructed Model	1.98	50.427	6723.70	127.78
Decomposition Model	1.55	38.46	2632.40	451.50
Ordinary Model	2.32	59.69	8898.00	41.00

Table 5. Comparison of the reconstructed model and the decomposition model.

5.5. Interval Prediction Based on Kernel Density Estimation

To better estimate the uncertainty in the load sequence, we used the kernel density estimation method to estimate the load interval. Firstly, we use the maximum real load value of the training set to normalize the error of the training set, and then divide the error into 0–1750 MW, 1750–2350 MW, 2350–2850 MW, and more than 2850 MW, according to the size of the predicted load value. The four intervals are respectively estimated by kernel density estimation and logistic estimation, and the optimal approximation curve is selected. Then, according to the predicted value of the prediction set, the corresponding error percentage is selected to obtain the final prediction interval.

It can be seen from Figure 4 that the fitting effect of kernel density estimation is better than that of logistic estimation in the process of estimating the set error of the 0–1750 MW interval. Further comparison with Figure 4b, it can be found that the normal kernel has a better fitting effect on the cumulative distribution function curve of the training set error, and the error range is [-1.44%, +2.1%] under the 90% confidence interval. Similarly, we found that the prediction effect of 1750–2350 MW Epanechnikov kernel is better through experiments, and the error range of 90% confidence interval is [-2.9%, +2.6%]. For the 2350–2850 MW load interval, Box kernel has a good fitting. The error range is [-3.3%, +4.1%] under 90% confidence interval. The Box kernel above 2850 MW has a good prediction effect, and the corresponding value range is [-3.21%, +3.98%].



Figure 4. 0–1750 MW interval training set error; (**a**) probability density function curve; (**b**) cumulative distribution function curve.

Finally, prediction intervals coverage probability (PICP) is 0.919 and prediction intervals normalized averaged width (PINAW) is 0.112. PCIP is 0.919, indicating that 91.9% of the load values in the test set fall within the prediction interval, and PCIP > interval confidence, which shows that the model in this paper has good prediction performance and can accurately estimate the load change. For PINAW, when the prediction interval width is certain, the larger the variation range of real load data, the smaller the PINAW, which also represents the better performance of the model. To avoid the impact of the highest point of annual load value (4952) on PINAW, we select the second highest point of forecast set

of annual load value (4952) on PINAW, we select the second highest point of forecast set value 3416 as the upper limit of load change, and the final PINAW is 0.112. This shows that the width of the prediction interval is within a reasonable range, and the model used in this paper does not obtain high coverage by unlimited increase of the width of the error interval. To sum up, we can conclude that the probability prediction model proposed in this paper has good prediction accuracy.

6. Conclusions

By analyzing the above experiments, we can draw the following conclusions:

- (1) Compared with EEMD and EMD decomposition models, we find that ICEMDAN decomposition has better prediction accuracy. In addition, through the comparison of the decomposition model, reconstruction model, and ordinary model, we can find that the reconstruction model performs well in training time and prediction accuracy, and is suitable for load forecasting scenarios. Combined ICEEMDAN with sample entropy is used to decompose and reconstruct the load series, which not only improves the accuracy of load forecasting, but also reduces the number of models, shortens the training time, and improves the forecasting efficiency.
- (2) Through the comparison between SSA-ELM and ELM, we can find that the prediction accuracy of the model has been significantly improved after using SSA to optimize the number and threshold of ELM hidden layer neurons. SSA-ELM can effectively improve the stability and accuracy of prediction results.
- (3) The kernel density estimation is used to analyze the error interval, which has a good fitting for the error curve and can obtain a more accurate prediction interval. We also found that the choice of different sum functions will affect the fitting effect of error distribution, and then affect the accuracy of interval prediction.
- (4) PICP was 0.919 and PINAW was 0.112. These two indicators show that the model achieves high coverage in a reasonable interval width. This means that the method used in this paper can better predict the variation range of load and reflect some unknown load information. It also proves the feasibility of the method used in this paper.

Author Contributions: Conceptualization, T.H. and M.Z.; methodology, T.H. and M.Z.; software, T.H. and K.B.; validation, T.H. and W.L.; writing—original draft preparation, T.H.; data curation, T.H. and Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China, grant number: 2018YFC0604503; the Energy Internet Joint Fund Project of Anhui province, grant number: 2008085UD06; the Major Science and Technology Program of Anhui Province, grant number: 201903a07020013; the Ministry of Education New Generation of Information Technology Innovation Project, grant number: 2019ITA01010; and the Demonstration Project of Science Popular-ization Innovation and Scientific Research Education for College Students, grant number: KYX202117.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from ENTSO-E and is available at https://transparency.entsoe.eu/dashboard/show (accessed on 15 October 2021).

Acknowledgments: We would like to thank the ENTSO-E for making the data available.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Wang, B.; Wang, Q.; Wei, Y.-M.; Li, Z.-P. Role of renewable energy in China's energy security and climate change mitigation: An index decomposition analysis. *Renew. Sustain. Energy Rev.* **2018**, *90*, 187–194. [CrossRef]
- Zhao, J.; Oh, U.; Lee, Y.; Park, J.; Choi, J. A Study on Reliability and Capacity Credit Evaluation of China Power System Con-sidering WTG with Multi Energy Storage Systems. J. Electr. Eng. Technol. 2021, 16, 2367–2378. [CrossRef]
- Cao, R.; Sun, D.; Wang, L.; Yan, Z.; Liu, W.; Wang, X.; Zhang, X. Enhancing solar–thermal–electric energy conversion based on m-PEGMA/GO synergistic phase change aerogels. J. Mater. Chem. A 2020, 8, 13207–13217. [CrossRef]
- Jahan, I.S.; Snasel, V.; Misak, S. Intelligent Systems for Power Load Forecasting: A Study Review. *Energies* 2020, 13, 6105. [CrossRef]
- 5. Borenius, S.; Hämmäinen, H.; Lehtonen, M.; Ahokangas, P. Smart grid evolution and mobile communications—Scenarios on the Finnish power grid. *Electr. Power Syst. Res.* **2021**, *199*, 107367. [CrossRef]
- 6. Li, X.; Dong, C.; Jiang, W.; Wu, X. Distributed control strategy for global economic operation and bus restorations in a hybrid AC/DC microgrid with interconnected subgrids. *Int. J. Electr. Power Energy Syst.* **2021**, *131*, 107032. [CrossRef]
- Sharma, S.; Majumdar, A.; Elvira, V.; Chouzenoux, E. Blind Kalman Filtering for Short-Term Load Forecasting. *IEEE Trans. Power* Syst. 2020, 35, 4916–4919. [CrossRef]
- 8. Singh, P.; Dwivedi, P.; Kant, V. A hybrid method based on neural network and improved environmental adaptation method using Controlled Gaussian Mutation with real parameter for short-term load forecasting. *Energy.* **2019**, *174*, 460–477. [CrossRef]
- 9. Zhang, W.; Quan, H.; Gandhi, O.; Rajagopal, R.; Tan, C.-W.; Srinivasan, D. Improving Probabilistic Load Forecasting Using Quantile Regression NN With Skip Connections. *IEEE Trans. Smart Grid* **2020**, *11*, 5442–5450. [CrossRef]
- 10. Gan, D.; Wang, Y.; Yang, S.; Kang, C. Embedding based quantile regression neural network for probabilistic load forecasting. *J. Mod. Power Syst. Clean Energy* **2018**, *6*, 244–254. [CrossRef]
- 11. Toubeau, J.J.; Bottieau, J.; Vallée, F.; De Grève, Z. Deep Learning-Based Multivariate Probabilistic Forecasting for Short-Term Scheduling in Power Markets. *IEEE Trans. Power Syst.* **2019**, *34*, 1203–1215. [CrossRef]
- 12. Brusaferri, A.; Matteucci, M.; Portolani, P.; Vitali, A. Bayesian deep learning based method for probabilistic forecast of day-ahead electricity prices. *Appl. Energy* **2019**, *250*, 1158–1175. [CrossRef]
- 13. Huang, Q.; Li, J.; Zhu, M. An improved convolutional neural network with load range discretization for probabilistic load forecasting. *Energy* **2020**, 203, 117902. [CrossRef]
- 14. Guan, J.; Lin, J.; Guan, J.; Mokaramian, E. A novel probabilistic short-term wind energy forecasting model based on an im-proved kernel density estimation. *Int. J. Hydrogen Energy* **2020**, *45*, 23791–23808. [CrossRef]
- 15. Memarzadeh, G.; Keynia, F. Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. *Electr. Power Syst. Res.* 2020, 192, 106995. [CrossRef]
- Kong, W.; Dong, Z.Y.; Jia, Y.; Hill, D.J.; Xu, Y.; Zhang, Y. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. *IEEE Trans. Smart Grid* 2017, 10, 841–851. [CrossRef]
- 17. Yin, L.; Xie, J. Multi-temporal-spatial-scale temporal convolution network for short-term load forecasting of power systems. *Appl. Energy* **2020**, *283*, 116328. [CrossRef]
- 18. Afrasiabi, M.; Mohammadi, M.; Rastegar, M.; Stankovic, L.; Afrasiabi, S.; Khazaei, M. Deep-Based Conditional Probability Density Function Forecasting of Residential Loads. *IEEE Trans. Smart Grid* **2020**, *11*, 3646–3657. [CrossRef]
- 19. Tan, Z.; Zhang, J.; He, Y.; Zhang, Y.; Xiong, G.; Liu, Y. Short-Term Load Forecasting Based on Integration of SVR and Stacking. *IEEE Access* 2020, *8*, 227719–227728. [CrossRef]
- Sina, A.; Kaur, D. Short Term Load Forecasting Model Based on Kernel-Support Vector Regression with Social Spider Optimization Algorithm. J. Electr. Eng. Technol. 2019, 15, 393–402. [CrossRef]
- 21. Xie, K.; Yi, H.; Hu, G.; Li, L.; Fan, Z. Short-term power load forecasting based on Elman neural network with particle swarm optimization. *Neurocomputing* **2019**, *416*, 136–142. [CrossRef]
- 22. Janković, Z.; Selakov, A.; Bekut, D.; Đorđević, M. Day similarity metric model for short-term load forecasting supported by PSO and artificial neural network. *Electr. Eng.* **2021**, *103*, 2973–2988. [CrossRef]
- 23. Zheng, H.; Yuan, J.; Chen, L. Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation. *Energies* 2017, *10*, 1168. [CrossRef]
- 24. Xiang, Y.; Gou, L.; He, L.; Xia, S.; Wang, W. A SVR–ANN combined model based on ensemble EMD for rainfall prediction. *Appl. Soft Comput.* **2018**, *73*, 874–883. [CrossRef]
- Ge, Q.; Guo, C.; Jiang, H.; Lu, Z.; Yao, G.; Zhang, J.; Hua, Q. Industrial Power Load Forecasting Method Based on Reinforcement Learning and PSO-LSSVM. *IEEE Trans. Cybern.* 2020, 99, 1–13. [CrossRef]
- 26. Zhang, Z.; Hong, W.-C. Electric load forecasting by complete ensemble empirical mode decomposition adaptive noise and support vector regression with quantum-based dragonfly algorithm. *Nonlinear Dyn.* **2019**, *98*, 1107–1136. [CrossRef]
- Rafi, S.H.; Nahid-Al-Masood; Deeba, S.R.; Hossain, E. A Short-Term Load Forecasting Method Using Integrated CNN and LSTM Network. *IEEE Access* 2021, 9, 32436–32448. [CrossRef]
- Wang, Y.; Zhang, N.; Chen, X. A Short-Term Residential Load Forecasting Model Based on LSTM Recurrent Neural Network Considering Weather Features. *Energies* 2021, 14, 2737. [CrossRef]

- 29. Phyo, P.P.; Jeenanunta, C. Daily Load Forecasting Based on a Combination of Classification and Regression Tree and Deep Belief Network. *IEEE Access* 2021, *9*, 152226–1522421. [CrossRef]
- Sharma, E.; Deo, R.C.; Prasad, R.; Parisi, A.V. A hybrid air quality early-warning framework: Hourly forecasting with online sequential extreme learning machine and empirical mode decomposition algorithm. *Sci. Total Environ.* 2019, 709, 135934. [CrossRef]
- Yang, Z.-X.; Wang, X.-B.; Wong, P.K. Single and Simultaneous Fault Diagnosis with Application to a Multistage Gearbox: A Versatile Dual-ELM Network Approach. *IEEE Trans. Ind. Inform.* 2018, 14, 5245–5255. [CrossRef]
- Lan, Y.; Hu, J.; Huang, J.; Niu, L.; Zeng, X.; Xiong, X.; Wu, B. Fault diagnosis on slipper abrasion of axial piston pump based on Extreme Learning Machine. *Measurement* 2018, 124, 378–385. [CrossRef]
- 33. Bian, K.; Zhou, M.; Hu, F.; Lai, W. RF-PCA: A New Solution for Rapid Identification of Breast Cancer Categorical Data Based on Attribute Selection and Feature Extraction. *Front. Genet.* **2020**, *11*, 566057. [CrossRef] [PubMed]
- 34. Colominas, M.A.; Schlotthauer, G.; Torres, M.E. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. *Biomed. Signal Processing Control* 2014, 14, 19–29. [CrossRef]
- 35. Richman, J.S.; Lake, D.E.; Moorman, J.R. Sample Entropy. Methods Enzymol. 2004, 384, 172–184.
- Mirjalili, S.; Gandomi, A.H.; Mirjalili, S.Z.; Saremi, S.; Faris, H.; Mirjalili, S.M. Salp swarm algorithm: A bio-inspired optimizer for engineering design problems. *Adv. Eng. Softw.* 2017, 114, 163–191. [CrossRef]
- Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing* 2006, 70, 489–501. [CrossRef]
- 38. Rosenblatt, M. Remarks on Some Nonparametric Estimates of a Density Function. *Ann. Math. Stat.* **1956**, *27*, 832–837. [CrossRef]
- 39. Parzen, E. On Estimation of a Probability Density Function and Mode. Ann. Math. Stat. 1962, 33, 1065–1076. [CrossRef]
- 40. Hu, B.; LI, Y.; Yang, H.; Wang, H. Wind speed model based on kernel density estimation and its application in reliability as-sessment of generating systems. *J. Mod. Power Syst. Clean Energy* **2017**, *5*, 220–227. [CrossRef]